Project Statement

BIGDATA SUMMER IMMERSION

Public Health Modeling Group

Instructors: Stephanie Perniciaro, PhD, MPH and Shelby Golden, MS

Public health is an interdisciplinary field that collectively measures and tracks disease trends in populations, whether of humans, animals, plants, etc., and promotes actions through policy or treatment interventions that promote overall health and wellbeing. Population health modeling uses quantitative tools to explore the biological, sociological, and environmental variables that contribute to the health of populations. In our project, we will model pneumococcal disease, one of the leading causes of deaths in children under 5 globally.

Pneumococcal disease is caused by the bacteria *Streptococcus pneumoniae*. This species is a particular oddity due to its non-disease-causing status in about half of children and one third of adults globally. In its asymptomatic state, it lives in the microflora of a person's nose and throat. It is when the bacteria migrate down to the lungs, up into the brain, or into the bloodstream that it becomes a major cause of disease. It is the most common cause of bacterial pneumonia globally and is a leading cause for infections ranging from earaches to sepsis.

Because of the different diseases pneumococci can cause, diagnosing and treating pneumococcal infections is complicated. Preventing pneumococcal infections before they happen has been a winning strategy. Several vaccines have been developed to prevent pneumococcal infections, but they are can only protect against certain types of pneumococci, also referred to as serotype.

When an effective vaccine is implemented in a population, some of the disease-causing pneumococci serotypes reduce or disappear altogether. Unfortunately, this allows for other pneumococci serotypes to flourish and take over as the predominant causes of disease. This phenomenon is called serotype replacement, and it has been documented in populations throughout the world after the implementation of effective pneumococcal vaccine programs.

Our objectives for this course will be to:

- Characterize and contextualize pneumococcal serotype replacement in several global populations.
- Explore the implications of different vaccine programs on the serotype distribution of invasive pneumococcal disease.
- Discuss biological, epidemiological, and policy choices that can contribute to our understanding of serotype replacement and pneumococcal evolution.

Version: 001.07.09.25

We will accomplish these objectives using data from national and multinational infectious disease surveillance systems, and the following statistical methods:

- Time series analysis
- Hierarchical modeling
- Sequence clustering with the Global Pneumococcal Sequencing (GPS) project analysis pipeline
- Association detection Market Basket

The analyses for this course will be completed in R.